Intravoxel incoherent motion: model-free determination of tissue type in abdominal organs using machine learning

Ciritsis, Alexander ; Rossi, Cristina ; Wurnig, Moritz C ; Phi Van, Valerie Doan ; Boss, Andreas

Abstract: PURPOSE: For diffusion data sets including low and high b-values, the intravoxel incoherent motion model is commonly applied to characterize tissue. The aim of the present study was to show that machine learning allows a model-free approach to determine tissue type without a priori assumptions on the underlying physiology. MATERIALS AND METHODS: In 8 healthy volunteers, diffusion data sets were acquired using an echo-planar imaging sequence with 16 b-values in the range between 0 and 1000 s/mm. Using the k-nearest neighbors technique, the machine learning algorithm was trained to distinguish abdominal organs (liver, kidney, spleen, muscle) using the signal intensities at different b-values as training features. For systematic variation of model complexity (number of neighbors), performance was assessed by calculation of the accuracy and the kappa coefficient ( ). Most important b-values for tissue discrimination were determined by principal component analysis. RESULTS: The optimal trade-off between model complexity and overfitting was found in the range between K = 11 to 13. On "real-world" data not previously applied to optimize the algorithm, the k-nearest neighbors algorithm was capable to accurately distinguish tissue types with best accuracy of 94.5% and = 0.92 reached for intermediate model complexity (K = 11). The principal component analysis showed that most important b-values are (with decreasing importance): b = 1000 s/mm, b = 970 s/mm, b = 750 s/mm, b = 20 s/mm, b = 620 s/mm, and b = 40 s/mm. Applying a reduced set of 6 most important b-values, still a similar accuracy was achieved on the real-world data set with an average accuracy of 93.7% and a coefficient of 0.91. CONCLUSIONS: Machine learning allows for a model-free determination of tissue type using intra voxel incoherent motion signal decay curves as features. The technique may be useful for segmentation of abdominal organs or distinction between healthy and pathological tissues.

DOI: https://doi.org/10.1097/RLI.0000000000000400

Originally published at:
Ciritsis, Alexander; Rossi, Cristina; Wurnig, Moritz C; Phi Van, Valerie Doan; Boss, Andreas (2017). Intravoxel incoherent motion: model-free determination of tissue type in abdominal organs using machine learning. Investigative Radiology, 52(12):747-757.
DOI: https://doi.org/10.1097/RLI.0000000000000400

Intravoxel Incoherent Motion
Model-Free Determination of Tissue Type in Abdominal Organs Using Machine Learning

Alexander Ciritsis, PhD, Cristina Rossi, PhD, Moritz C. Wurnig, MSc, MD, Valerie Phi Van, MD, and Andreas Boss, MD, PhD

Purpose: For diffusion data sets including low and high b-values, the intravoxel incoherent motion model is commonly applied to characterize tissue. The aim of the present study was to show that machine learning allows a model-free approach to determine tissue type without a priori assumptions on the underlying physiology.

Materials and Methods: In 8 healthy volunteers, diffusion data sets were acquired using an echo-planar imaging sequence with 16 b-values in the range between 0 and 1000 s/mm^2. Using the k-nearest neighbors technique, the machine learning algorithm was trained to distinguish abdominal organs (liver, kidney, spleen, muscle) using the signal intensities at different b-values as training features. For systematic variation of model complexity (number of neighbors), performance was assessed by calculation of the accuracy and the kappa coefficient (κ). Most important b-values for tissue discrimination were determined by principal component analysis.

Results: The optimal trade-off between model complexity and overfitting was found in the range between K = 11 to 13. On "real-world" data not previously applied to optimize the algorithm, the k-nearest neighbors algorithm was capable to accurately distinguish tissue types with best accuracy of 94.5% and $\kappa = 0.92$ reached for intermediate model complexity (K = 11). The principal component analysis showed that most important b-values are (with decreasing importance): b = 1000 s/mm^2, b = 970 s/mm^2, b = 750 s/mm^2, b = 20 s/mm^2, b = 620 s/mm^2, and b = 40 s/mm^2. Applying a reduced set of 6 most important b-values, still a similar accuracy was achieved on the real-world data set with an average accuracy of 93.7% and a κ coefficient of 0.91.

Conclusions: Machine learning allows for a model-free determination of tissue type using intra voxel incoherent motion signal decay curves as features. The technique may be useful for segmentation of abdominal organs or distinction between healthy and pathological tissues.

Key Words: machine learning, deep learning, IVIM, tissue classification

Invest Radiol 2017;52: 00–00)

Diffusion-weighted imaging (DWI) is a functional magnetic resonance imaging (MRI) technique, which initially was developed for brain-related applications such as stroke diagnosis. However, in recent years, it has also been widely implemented in clinical MRI protocols for examination of the breast, the liver, the prostate, and the rectum.[1–6] By obtaining the image contrast from the differences in mobility of water protons, DWI can be used to identify potential areas of impeded diffusion and has turned out as an important tool for tumor characterization, treatment monitoring, or defining parenchymal diseases.[7–9]

Received for publication April 19, 2017; and accepted for publication, after revision, June 15, 2017.

From the *Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Zurich, Switzerland.

Conflicts of interest and sources of funding: none declared.

Correspondence to: Alexander Ciritsis, PhD, Institute of Diagnostic and Interventional Radiology, University Hospital Zurich, Rämistr. 100 8091, Zurich, Switzerland. E-mail: alexander.ciritsis@usz.ch.

ISSN: 0020-9996/17/5212–0000
DOI: 10.1097/RLI.0000000000000400

The apparent diffusion coefficient is used to describe the degree of water diffusion. For calculation of the apparent diffusion coefficient, a monoexponential decay dependent on the applied b-value (a sequence parameter describing the strength of the diffusion weighting) is assumed. However, this classical monoexponential model does not reflect the true nature of water mobility in tissues because parts of the signal decay result from perfusion effects especially at low b-values (<150 s/mm^2). To take this "pseudodiffusion" effect caused by fast-moving water molecules in capillaries into account, the concept of intra voxel incoherent motion (IVIM) was introduced.[10,11] Based on this extended model, the overlapping contributions of normal diffusion and pseudodiffusion can be separated and quantified by fitting a biexponential signal curve to the measured signal intensities at a large number of b-values (typically in the order 12–16).

However, despite a broad ongoing research in the field of IVIM,[12–17] there are several shortcomings to be considered. First, there is still no consensus reached on the optimal number and resolution of applied b-values.[18–20] Second, the more complex fitting procedure of IVIM data due to the larger number of fitting parameters causes instability of calculated IVIM parameters, especially when it comes to the pseudodiffusion coefficient D*.[12] Third, for all IVIM approaches, a priori assumptions on the underlying physiology are required without knowing to which extent the chosen model describes reality and is stable in the presence of pathological tissue conditions.

In the present investigation, we propose and test a new approach overcoming the forementioned shortcomings: the tissue is characterized (classified) using a machine learning algorithm without a priori assumptions in a model-free approach. The purpose of this study was to train a k-nearest neighbor (K-NN) machine learning algorithm to classify different organs in the abdomen based on IVIM data and to determine the optimal combination of b-values for the distinction of the different organs using cluster analysis and principal component analysis (PCA).

MATERIALS AND METHODS

Volunteers

Eight healthy volunteers (6 men, 2 women; age range, 23–39 years; median age, 26 years) were included in this prospective study. Written informed consent was obtained from all volunteers, and the study was approved by the local ethics committee.

MRI Acquisition

Magnetic resonance imaging examinations were performed in supine position on a clinical 3.0-T scanner (Ingenia; Phillips, Best, the Netherlands) using a flexible 16-channel body matrix-coil and the built-in spine coil for signal reception. The sequence protocol included an echo-planar imaging (EPI) sequence to acquire DWI data sets of the upper abdomen in transverse orientation during free breathing. As suggested by Lemke et al,[19] the following 16 b-values were chosen for the DWI sequence (0, 10, 20, 40, 90, 100, 170, 200, 210, 240, 390,

530, 620, 750, 970, and 1000 s/mm^2). Additional sequence parameters were as follows: repetition time, 5000 milliseconds; echo time, 69.5 milliseconds; matrix size, 144 × 144; in-plane resolution, 2.8 × 2.8 mm; slice thickness, 5 mm; number of averages, 6; parallel imaging SENSE factor, 2; spectral attenuated inversion recovery fat suppression; receiver bandwidth, 2321 Hz/px.

Region of Interest Definition

A region of interest (ROI) analysis was performed to generate signal intensity curves from the DWI data sets as a function of the b-value for abdominal organs. All ROIs were defined using institutional computer scripts written in the programming language Matlab (The MathWorks, Natick, MA). Polygonal ROIs were defined on the b = 0 s/mm^2 image and subsequently copied to all other b-value images of the same data set, with an average ROI size of 154 pixels ranging between 70 and 200 pixels (Fig. 1 shows typical definitions of ROIs). For the following body regions, signal curves were assessed: liver; spleen, left kidney, right kidney, and erector spinae muscle. To acquire a sufficient set of data points for training and validation of the machine learning algorithm, a number of 10 ROIs were placed on different slices for each body region in all subjects resulting in a total number of 400 ROIs. For all ROIs, large vascular structures were avoided to exclude contributions from macroscopic vascular flow and to assess only parenchyma of the examined organ. The acquired signal intensity curves were normalized to 1 for the b = 0 s/mm^2 image.

IVIM Analysis

In the standard IVIM model, the signal intensity curves from a DWI experiment with multiple b-values are described by the following equation:

$$\frac{S_b}{S_0} = F_p \times \exp(-bD^*) + \left(1 - F_p\right) \times \exp(-\mathrm{b}D_t) \qquad (1)$$

with S_b being the signal intensity in the DWI image acquired with the respective b-value and S_0 representing the signal intensity in the b = 0 s/mm^2 image. F_p is the perfusion fraction, D^* standing for the pseudodiffusion coefficient representing capillary microperfusion, and D_t representing the "true diffusion" coefficient of passive molecular water diffusion. For the fitting of Equation 1 to the obtained signal intensities, an adaptive b-value threshold algorithm was used as described in a previous study by Wurnig et al.[12]

Machine Learning

Creating the Model

For the current study, a machine learning model based on the K-NN algorithm[21] was implemented to classify the different tissue types of abdominal organs and skeletal muscle. The signal intensity curves as functions of the b-value for each subject and organ obtained by ROI analyses were imported into the data-mining program Rapidminer (RapidMiner Inc, Boston, MA). The classification in the K-NN model was done by determining the k closest training examples to the test data set, whereas "closeness" is defined in terms of a distance metric, which in our study was the Euclidean distance (Equation 2).

$$\mathrm{D}(x, y) = \sqrt{\sum_{i=1}^{K} (x_i - y_i)^2} \qquad (2)$$

with D being the distance between 1 point in the training data set (x_i) and 1 point in the test data set (y_i) with 2 coordinates for each data point: the signal intensity S_b and the corresponding b-value.

Data Split

The imported data were split by a split data operator, dividing the data set randomly stratified in 2 partitions with a ratio of 0.7 to 0.3. The partition with 70% of the data served to train and validate the model. The partition with 30% of the data was excluded from the training and validation phase, and spared as testing data never "seen" by the machine. After implementation of the model, the testing data were evaluated by the model to measure the unbiased performance of the machine learning algorithm. In the following, this 30% data subset will be referred to as "real-world" data set, as it represents a real-world scenario with data, which did not serve to train or validate the machine learning algorithm.

Training and Validating the Model

To increase the data size, a 10-fold cross-validation operand consisting of a training and a validation component was used to split the data into a training and validation subset. The partition of 70% data was split into n = 10 subsets of equal size via stratified shuffled sampling. For each of the computation cycles, 1 single set of the 10 subsets was retained as testing data set, whereas the remaining 9 subsets were used to train the K-NN–based model. The training/validation process used in this study is described in Figure 2.

Performance of the Model

To evaluate and measure the overall performance as well as the individual accuracy of tissue classification of the implemented K-NN–based model, confusion matrices for systematic variation of model complexity (numbers of K closest neighbors: 3, 5, 7, 9, 11, 13) were generated. The overall and individual performances of tissue classification of the implemented K-NN–based model were measured on the cross-validated training data set as well as on the real-world data set.

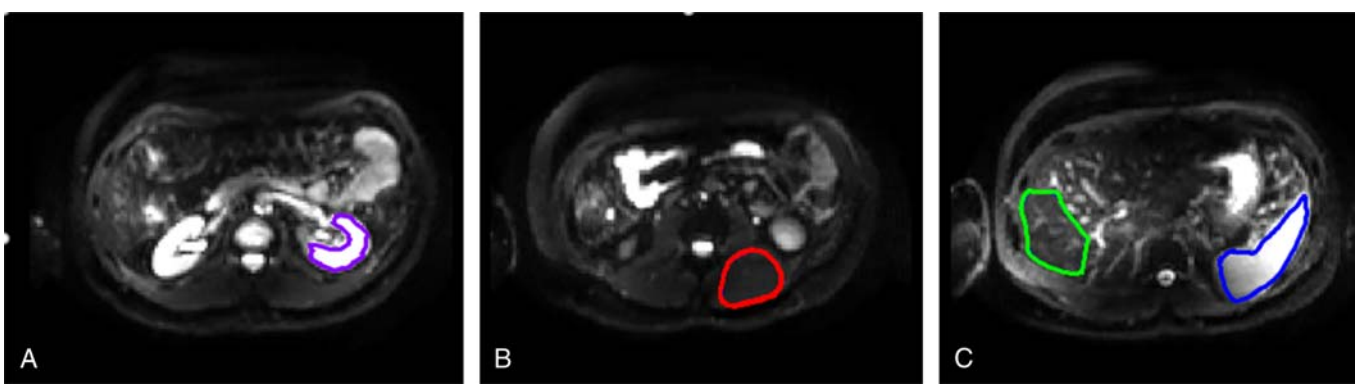

FIGURE 1. ROI analysis on the diffusion-weighted MRI acquired at a b-value of 0 s/mm^2: renal parenchyma (A), erector spinae muscle (B), and liver parenchyma splenic parenchyma (C). Figure 1 can be viewed online in color at www.investigativeradiology.com.

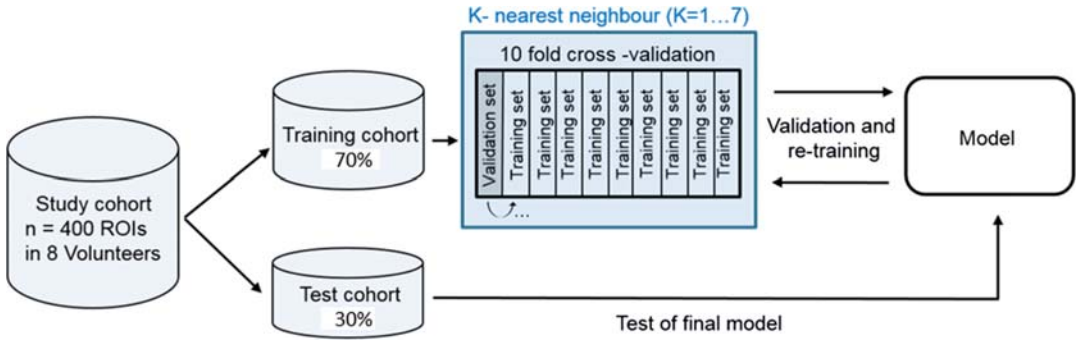

FIGURE 2. Schematic of the training/validation and test process used in this study. Figure 2 can be viewed online in color at www.investigativeradiology.com.

Overfitting and Underfitting

To find the "sweet spot" as the optimum trade-off between model complexity and training and test accuracy, the average classification error of the implemented K-NN algorithm was measured on the training as well as on the real-world data set. The average classification error was determined for 19 different values of K, declining from 100 to 1 with K = 100 representing a very low complexity of the model and K = 1, a high complexity, respectively. The determined error rates on each data set were plotted against the according reciprocal value of K (Figs. 3A, B).

Importance of Individual b-Values

Scatter Plots

To evaluate the importance of individual b-values for correct tissue classification, the distributions of the signal intensities curves for all b-values and each tissue type were assessed. Subsequent, scatter plots of the 4 tissue types for each b-value were generated and the 6 b-values, which separated the different tissue types best, were chosen visually.

Principal Component Analysis

To obtain a quantifiable value of the importance of different b-values, a PCA was implemented on the training data set. The conventional PCA is a mathematical procedure, using an orthogonal transformation, which aims to reduce a multivariate high dimensional correlated data set (1 axis per variable) to a lower uncorrelated dimensional subspace of principal components (PCs) (1 axis per PC). Thereby, the first PC accounts for as much of the variability in the data set as possible. The second PC is orthogonal to the first one and contains the second highest value of variance of the observed variables to describe the data. This procedure can be continued to as much PCs as variables are available. In terms of our imaging data, this meant that each PC was composed by a linear combination of all acquired b-values and their corresponding signal intensities in the different tissue types. The relevance of each b-value for the determined PCs is described by its eigenvector and can be visualized in a biplot, whereas the angle of the eigenvector with respect to the PC axis represents the contribution, and the amount of the eigenvector represents the weight of the variable to the corresponding PC. In our study, the 6 b-values, which contributed most on the 2 PCs to classify the different tissue types, were determined via PCA.

Performance of the Model With a Reduced Set of b-Values

On the reduced set of most contributing b-values, the overall performance of the implemented K-NN–based model was measured and evaluated. Correspondingly to the previous determined performances with the complete set of b-values, confusion matrices for systematic variation of model complexity (numbers of K closest neighbors: 3, 5, 7, 9, 11, 13) were generated for the reduced b-value set.

Statistical Evaluation

With the generated confusion matrices, the overall accuracy of the implemented K-NN algorithm was determined for each value of K by summing the number of correctly classified tissue types in the main diagonal and dividing by the total number of counts. Furthermore, to measure the agreement between classification and true counts, the kappa coefficient (κ) was determined, whereas a κ value of 1 stands for perfect agreement, whereas a value of 0 represents no agreement.

RESULTS

IVIM Analysis

All DW images were acquired successfully, and ROI analysis was performed for all volunteers. Typical examples of signal decay curves for the evaluated parenchymal abdominal organs including the corresponding biexponential IVIM fit are depicted for 1 volunteer in Figure 4A and as mean values for the whole cohort in Figure 4B. In both cases, the typical IVIM behavior of the signal curves can be seen with rapid signal decay for low b-values, especially for liver and spleen indicating a large contribution of perfusion effects. The corresponding IVIM parameters (pseudodiffusion D^*, perfusion fraction F_p, and true diffusion D_t) are reported in Figure 5 for the example volunteer and in Figure 6 for the whole cohort.

Model Complexity for 16 b-Values

For a value of K = 100 and associated low model complexity, the overall performance of the machine learning algorithm revealed a high average classification error of 49.7% and 41.8% on the training data set and real-world data set, respectively. With higher model complexity and declining K, the overall performance on both data sets improved and the average classification error decreased. The sweet spot describing a low and almost similar classification error on both data sets was determined between K = 11 and K = 13. For lower values of K down to 1 and associated high model complexity, the average classification error on the testing data set increased again, while it decreased on the training data set, indicating overfitting of the applied model (Fig. 3A).

K-Nearest Neighbor Algorithm

The overall performance and the κ coefficient of the implemented K-NN model were successfully measured on the training data set as well as on the real-world data set for values of K ranging from 3 to 13, whereas K = 3 described a high model complexity and K = 13 a low complexity, respectively. The overall accuracy and different classification rates for each tissue type are summarized in Tables 1 and 2 for K = 3 to 13. The optimal trade-off between model

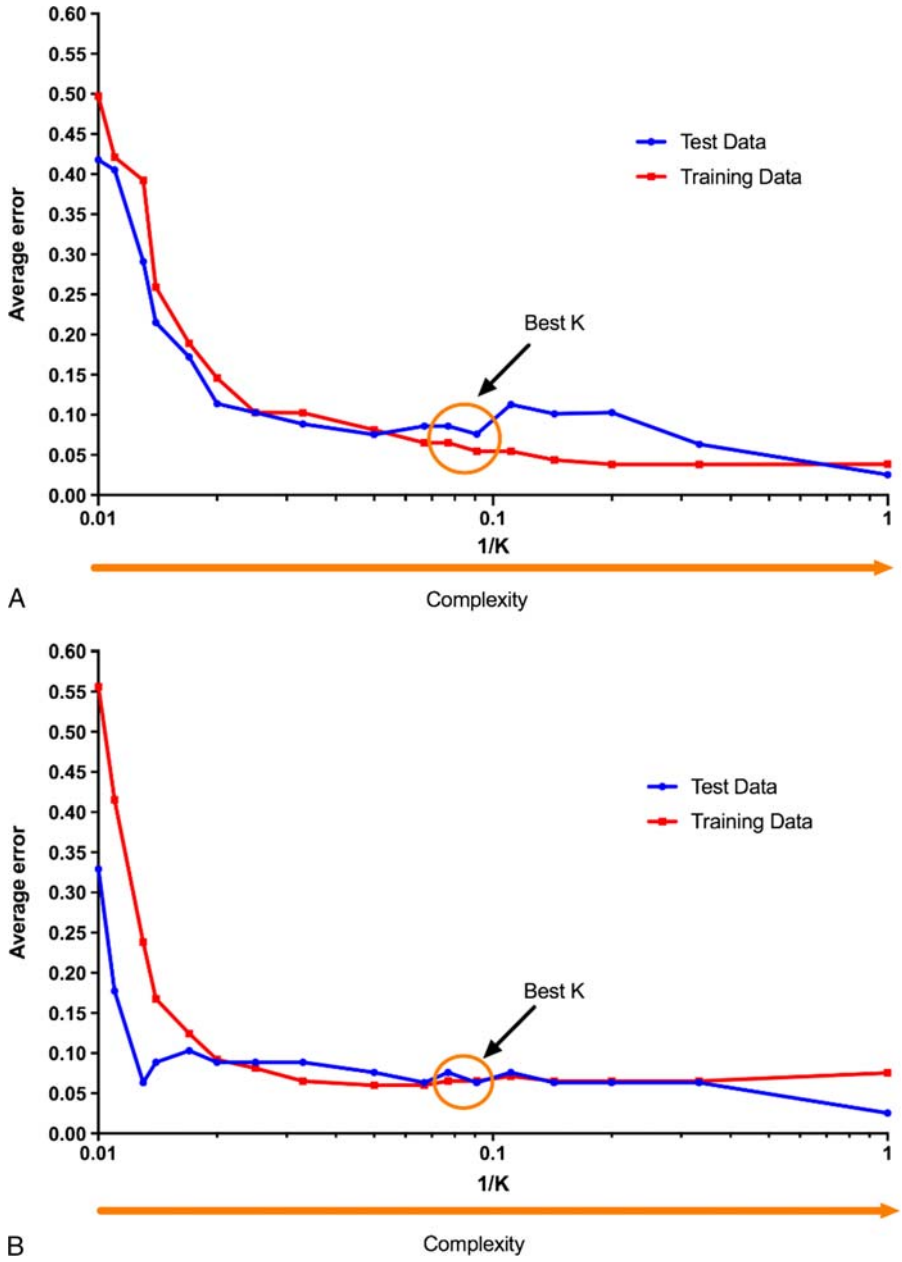

FIGURE 3. Performance for the K-NN model on the training and testing data sets, applying 16 (A) and 6 b-values (B) as attributes. For both scenarios, the average error decreases as the complexity of the model increases with declining K. For the complete set of 16 b-values, an optimal K between 11 and 13 was determined (A), as well as for the reduced data set of 6 b-values (B). Figure 3 can be viewed online in color at www.investigativeradiology.com.

complexity and overfitting was determined for a value of K ranging between 11 and 13, leading to similar performances of the model on the training and testing data set and thus minimizing overfitting. The confusion matrices of the training and real-world data set for K = 11 and K = 13 are exemplary given in Tables 3 and 4.

Intermediate Low Model Complexity (K = 11)

The overall accuracy for tissue classification on the training data set for K = 11 was 94.5% with $\kappa = 0.92$. Splenic and hepatic parenchyma reached the lowest correct classification rates of 89.2% and 91.7%, respectively, with splenic parenchyma being falsely classified as hepatic parenchyma and vice versa. Renal parenchyma and muscle tissue were mostly correctly classified with 98.6% and 95.2%, respectively (Table 3). On the real-world data set, a similar overall average accuracy of 92.4% was determined as on the training data set and the κ coefficient amounted 0.89. For muscle tissue splenic and also for hepatic parenchyma, the correct classification rate increased to 100% and 93.3%, whereas the classification rate of renal and splenic parenchyma decreased to 90% and 87.5% (Table 3). Also for the real-world data set, the most false classifications occurred between splenic and hepatic parenchyma.

Low Model Complexity (K = 13)

Compared with K = 11, the overall accuracy on the training data set measured for K = 13 slightly decreased to 93.5% with $\kappa = 0.91$. As for K = 11, splenic parenchyma reached the lowest correct classification

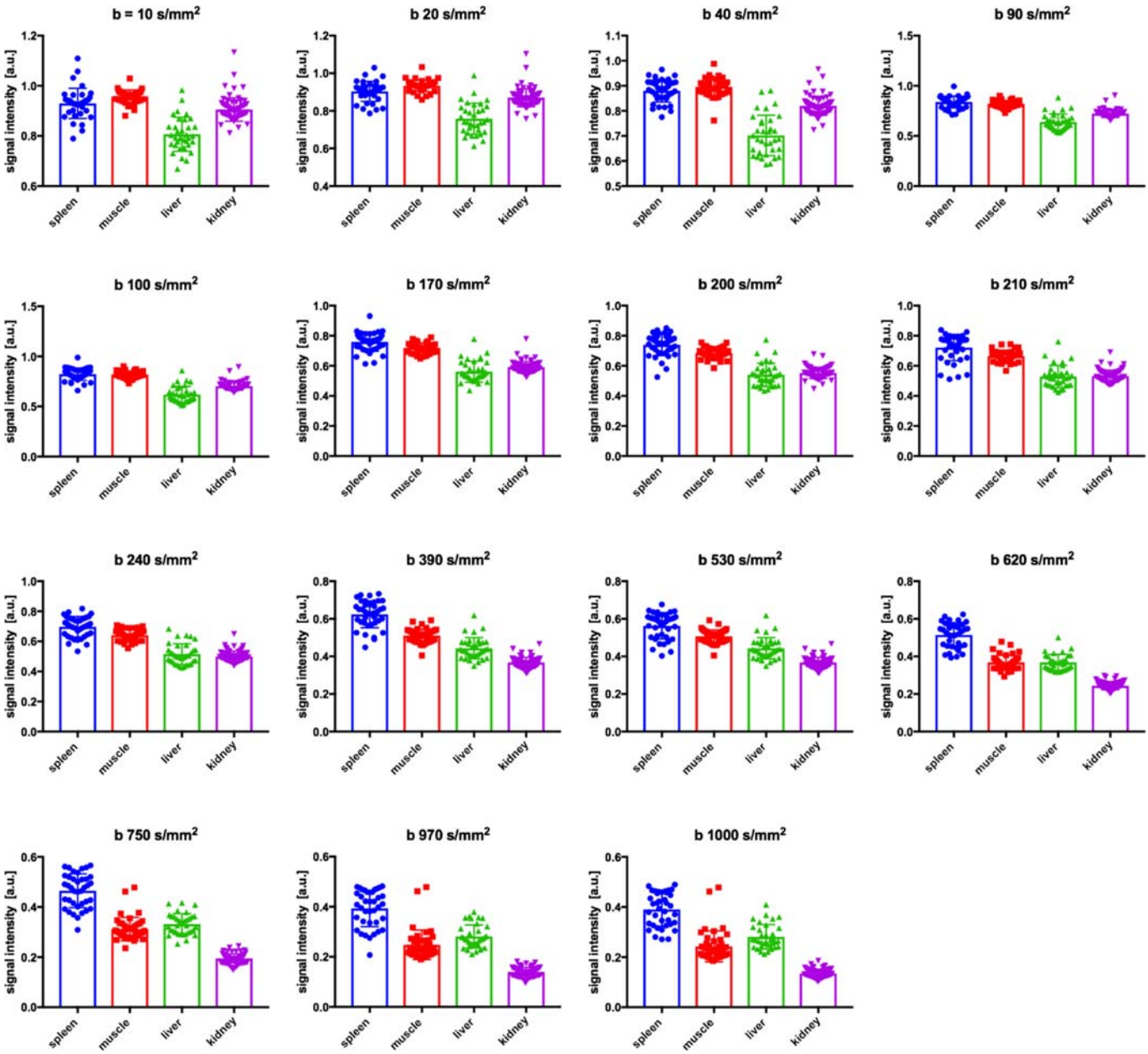

FIGURE 4. Scatter plots of 15 b-values, to classify the 4 different tissue types. With b-values in the low spectrum (b = 10 s/mm^2 to b = 90 s/mm^2), a separation of hepatic and splenic parenchyma is feasible. B-values in the high spectrum (b = 620 s/mm^2 to b = 1000 s/mm^2) contributed most to separate splenic from renal parenchyma and from muscle tissue. Figure 4 can be viewed online in color at www.investigativeradiology.com.

rate of 83.8% with almost all false classifications being assigned to hepatic parenchyma. The classification rate of hepatic parenchyma was similar to K = 11 with 91.7%. Muscle tissue and renal parenchyma again reached high correct classification rates of 95.2% and 98.5% (Table 4). On the real-world data set, the average overall accuracy increased to 91.1% with a corresponding κ coefficient of 0.88, with muscle tissue being 100% correctly classified and renal, splenic, and hepatic parenchyma reaching classification rates below 95% (Table 4).

Importance of Individual b-Values

The qualitative visual analysis of the scatter plots (Fig. 4) derived from the training data set revealed that b-values in the low spectrum, ranging from b = 10 s/mm^2 to b = 90 s/mm^2, contributed most to separate hepatic and splenic parenchyma from renal parenchyma and muscle tissue. For high b-values ranging from b = 620 s/mm^2 to b = 1000 s/mm^2, a separation of splenic to renal parenchyma and to muscle tissue was feasible. The b-values in the midrange ranging from b = 100 s/mm^2 to b = 530 s/mm^2 did not add much information regarding the separation of the different tissue types.

Principal Component Analysis

The PCA was implemented on the training data set, and the cumulative variance was set to 95%, meaning that all components greater than that given variance threshold were removed. For the first principal component (PC1), a cumulative variance of 75.9% was

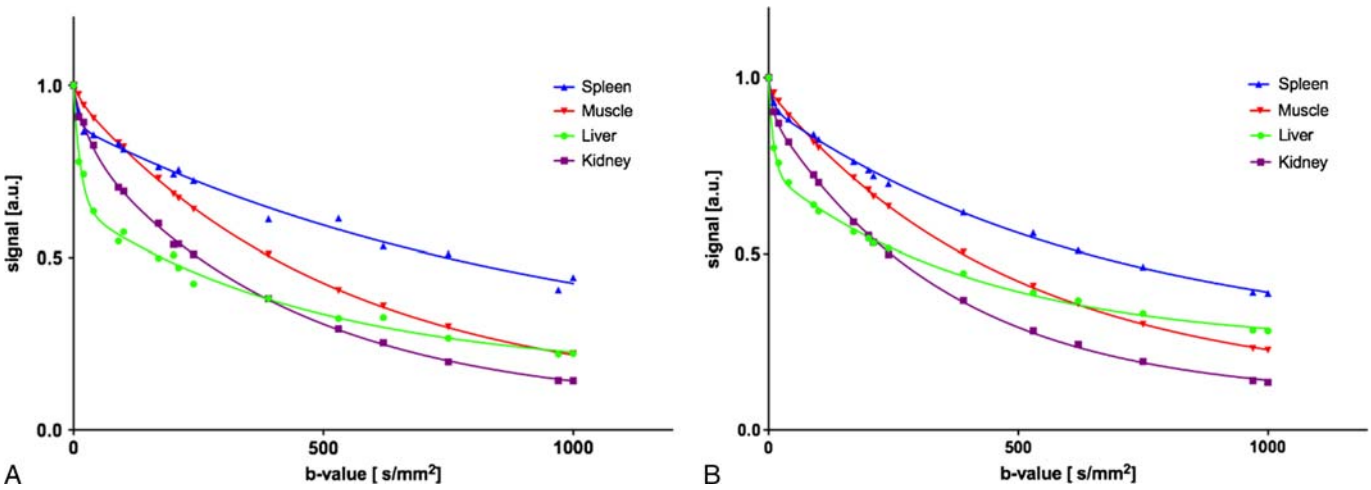

FIGURE 5. Example of the biexponential fitting curves for increasing b values in 4 different tissue types derived by ROI measurements for 1 volunteer (A) and for the whole cohort (B). Figure 5 can be viewed online in color at www.investigativeradiology.com.

determined, and for the second one (PC2), 95.8% was determined. Because the increase of variance for the other components was only minimal and the threshold was already achieved after the second PC, no further components were needed. A scatter plot subdividing the different tissue types based on the dimensions of the PCA was generated (Figs. 7, 8), showing a clear separation of all 4 tissue types from each other. Hepatic parenchyma and muscle tissue were mainly separated due to the influence of the second PC (*y*-axis), whereas the first PC (*x*-axis) had major influence on splenic and renal parenchyma. The importance of each b-value for PC1 and PC2 is shown in Table 5, and visualized in a biplot (Fig. 8), thereby b = 1000 s/mm^2, b = 970 s/mm^2, b = 750 s/mm^2, b = 20 s/mm^2, b = 620 s/mm^2, and b = 40 s/mm^2 revealed as most significant with eigenvector lengths ranging from 0.381 to 0.425.

Model Complexity for Reduced Set of b-Values

Similar to the complete set of b-values, an optimum of K = 11 to 13 was determined for the reduced set of b-values (Fig. 3B). The overall accuracy and different classification rates for each tissue type are summarized in Table 2 for K = 3 to 13. With the confusion matrices of the training and real-world data set for K = 11 and 13, describing the sweet spot between model complexity and overfitting are exemplary given in Tables 3 and 4 (brackets). For intermediate complexity (K = 11), the overall accuracy for tissue classification on the training data set was 93.7% with $\kappa = 0.91$. On the real-world data set, the average accuracy was 93.7% with a κ coefficient of 0.90. For low model complexity (K = 13), the overall accuracy on the training data set measured 93.5% with $\kappa = 0.90$ (Table 4 brackets). On the real-world data set, the average accuracy decreased to 92.4% with κ being 0.89.

DISCUSSION

In the present study, we propose a model-free approach for the evaluation of IVIM data using machine learning with the signal intensities at different b-values serving as features. To show the feasibility of the approach, a relatively easy-to-understand K-NN machine learning algorithm, which classified 4 different tissue types in the abdomen based on IVIM-DWI data, was implemented and evaluated. We showed that an accurate classification of tissue type using IVIM data sets with 16 different b-values as attributes is feasible. Moreover, we determined the most important sequence of b-values for tissue discrimination using PCA, and we demonstrated that, by using only the 6 most important b-values, an almost identical accuracy of tissue discrimination can be achieved compared with the full data set. To our knowledge, this is the first time a machine learning algorithm was applied to classify different tissue types in the abdomen based on IVIM data.

In recent years, several studies dealt with the applicability of the IVIM concept for tissue and lesion characterization in abdominal organs.[9,22,23] One limitation all those studies had in common was the need of choosing a suitable pharmacokinetic or compartment model of the organ or tissue type to be characterized to implement the IVIM

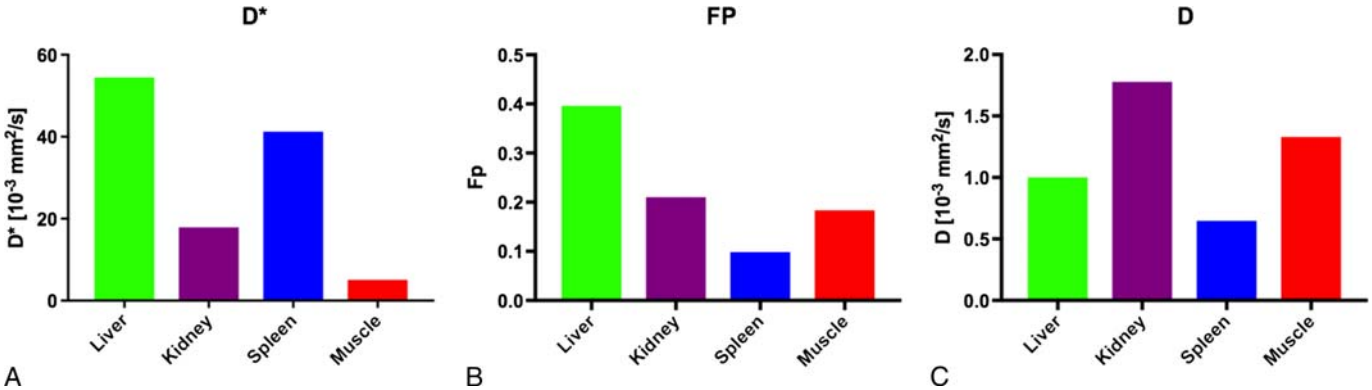

FIGURE 6. IVIM parameters. A, Pseudodiffusion D*. B, Perfusion fraction Fp. C, True diffusion D for 1 exemplary volunteer. Figure 6 can be viewed online in color at www.investigativeradiology.com.

TABLE 1. Overall Average Classification Accuracies and κ of the Implemented K-NN (K = 3 to 13) Algorithm Performed on the Training and Real-World Data Set for 16 and 6 b-Values

	16 b-Values		6 b-Values	
K-NN	Overall Accuracy Training Data Set, %	Overall Accuracy Real-World Data Set, %	Overall Accuracy Training Data Set, %	Overall Accuracy Real-World Data Set, %
3	96.17	93.67	93.48	93.67
	κ = 0.947	κ = 0.913	κ = 0.910	κ = 0.913
5	96.17	89.87	93.48	93.67
	κ = 0.947	κ = 0.862	κ = 0.910	κ = 0.913
7	95.61	89.87	93.48	93.67
	κ = 0.940	κ = 0.862	κ = 0.910	κ = 0.913
9	94.53	89.87	92.89	92.41
	κ = 0.924	κ = 0.862	κ = 0.901	κ = 0.896
11	94.53	92.41	93.71	93.67
	κ = 0.924	κ = 0.896	κ = 0.916	κ = 0.914
13	93.48	91.14	93.45	92.41
	κ = 0.910	κ = 0.879	κ = 0.909	κ = 0.896

K-NN, k-nearest neighbor.

analysis. Moreover, it is well known that model-based IVIM analysis often provides instable results (especially regarding the perfusion-dependent parameters), and numerous different approaches have been investigated to increase stability. With a machine learning algorithm only taking into account the signal intensities at different b-values, we implemented a completely model-free approach to classify different tissue types without any a priori knowledge or assumption of a model.

Applying techniques of machine learning, optimization regarding model complexity has to be done carefully ruling out overfitting. Overfitting occurs when the machine learning algorithm is trained in a too large extent with details and noise negatively affecting the performance on real-world data. To minimize overfitting in our study, the K-NN operator was trained via cross-validation, dividing the training data into 10 subsets of equal size. This resampling technique also provides advantages with respect to the limited number of data points. By holding back a subset of 30% data as validation subset for a real-world scenario, it was possible to estimate the true performance of the applied algorithm and to determine and validate overfitting or underfitting. Setting the value for K = 11 to 13, the overall accuracy on the training data showed an acceptable performance achieving 93% to 94% accuracy. In addition, the performance on the real-world data set also measured a high accuracy by classifying averagely 91.5% of the data correctly. Consequently, a value of K from 11 to 13 showed to be the best compromise between overall accuracy and overfitting.

The analysis of the IVIM data via machine learning allowed determining the most important b-values to distinguish tissue types. To identify the most relevant b-values for tissue classification, the distribution of the different tissue types depending on the b-values was visually assessed via scatter plots on the training data set. The analysis showed that b-values in the low spectrum ranging from b = 10 s/mm^2 to b = 90 s/mm^2 were most important to separate hepatic from splenic, renal, and muscle tissue. This finding was in accordance

TABLE 2. Individual Classification Rate for Each Tissue Type of the Implemented K-NN (K = 3 to 13) Algorithm Performed on the Training and Real-World Data Set (Brackets) for 16 and 6 b-Values

Training Data set (Real-World Data Set)	16 b-Values Accuracy				6 b-Values Accuracy			
K-NN	Spleen, %	Kidney, %	Muscle, %	Liver, %	Spleen, %	Kidney, %	Muscle, %	Liver, %
3	94.59	98.57	100	88.88	86.49	100	95.24	86.11
	(87.50)	(93.33)	(100)	(93.33)	(87.50)	(93.33)	(100.00)	(93.33)
5	91.89	98.57	97.62	94.44	83.78	100	95.24	88.89
	(75.00)	(90.00)	(100)	(93.33)	(93.75)	(90.00)	(100)	(93.33)
7	89.19	98.57	100.00	91.67	89.19	100	92.86	86.11
	(75.00)	(90.00)	(100.00)	(93.33)	(93.75)	(90.00)	(100.00)	(93.33)
9	86.49	98.57	97.62	91.67	89.19	100	92.86	83.33
	(75.00)	(90.00)	(100.00)	(93.33)	(87.50)	(90.00)	(100.00)	(93.33)
11	89.19	98.57	95.24	91.67	89.19	100	92.86	86.11
	(87.50)	(90.00)	(100.00)	(93.33)	(93.75)	(90.00)	(100.00)	(93.33)
13	83.78	98.57	95.24	91.67	89.19	100	92.86	86.11
	(81.25)	(90.00)	(100.00)	(93.33)	(87.50)	(90.00)	(100.00)	(93.33)

K-NN, k-nearest neighbor.

TABLE 3. Confusion Matrix for the Training and Real-World Data Set of 16 b-Values and 6 b-Values (Brackets) Derived by the K-NN Machine Learning Algorithm With a Value of K = 11

	True Spleen	True Kidney	True Muscle	True Liver	Class Precision
K = 11 training data set					
Predicted spleen	33	0	0	3	91.89%
	(33)	(0)	(1)	(5)	(84.62%)
Predicted kidney	0	69	2	0	98.61%
	(0)	(70)	(2)	(0)	(97.22%)
Predicted muscle	0	1	40	0	95.12%
	(0)	(0)	(39)	(0)	(100.00%)
Predicted liver	4	0	0	33	87.57%
	(4)	(0)	(0)	(31)	(88.57%)
Class recall	89.19%	98.57%	95.24%	91.67%	Average accuracy: 94.53%
	(89.19%)	(100%)	(92.86%)	(86.11%)	(93.71%)
K = 11 real-world data set					
Predicted spleen	14	0	0	1	93.33%
	(15)	(0)	(0)	(1)	(93.75%)
Predicted kidney	0	27	0	0	100%
	(0)	(27)	(0)	(0)	(100.00%)
Predicted muscle	0	3	18	0	90.91%
	(0)	(3)	(18)	(0)	(85.71%)
Predicted liver	2	0	0	14	94.12%
	(1)	(0)	(0)	(14)	(93.33%)
Class recall	87.50%	90.00%	100%	93.33%	Average accuracy: 92.00%
	(93.75%)	(90.00%)	(100.00%)	(93.33%)	(93.67%)

K-NN, k-nearest neighbor.

TABLE 4. Confusion Matrix for the Training and Real-World Data Set of 16 b-Values and 6 b-Values (Brackets) Derived by the K-NN Machine Learning Algorithm With a Value of K = 13

	True Spleen	True Kidney	True Muscle	True Liver	Class Precision
K = 13 training data set					
Predicted spleen	31	0	0	3	91.18%
	(33)	(0)	(1)	(5)	(84.62%)
Predicted kidney	0	69	2	0	97.18%
	(0)	(70)	(2)	(0)	(97.22%)
Predicted muscle	2	1	40	1	93.02%
	(0)	(0)	(39)	(1)	(100.00%)
Predicted liver	4	0	0	33	89.19%
	(4)	(0)	(0)	(31)	(88.57%)
Class recall	83.78%	98.57%	95.24%	91.67%	Average accuracy, 93.48%
	(89.19%)	(100.00%)	(92.86%)	(86.11%)	(93.45%)
K = 13 real-world data set					
Predicted spleen	13	0	0	1	92.89%
	(14)	(0)	(0)	(1)	(81.25%)
Predicted kidney	0	27	0	0	100%
	(0)	(27)	(1)	(0)	(96.3%)
Predicted muscle	0	3	18	0	85.71%
	(1)	(3)	(18)	(0)	(90.48%)
Predicted liver	3	0	0	14	82.35%
	(2)	(0)	(0)	(14)	(93.33%)
Class recall	81.25%	90.00%	100%	93.33%	Average accuracy, 91.14%
	(87.50%)	(90.00%)	(100.00%)	(93.33%)	(92.41%)

K-NN, k-nearest neighbor.

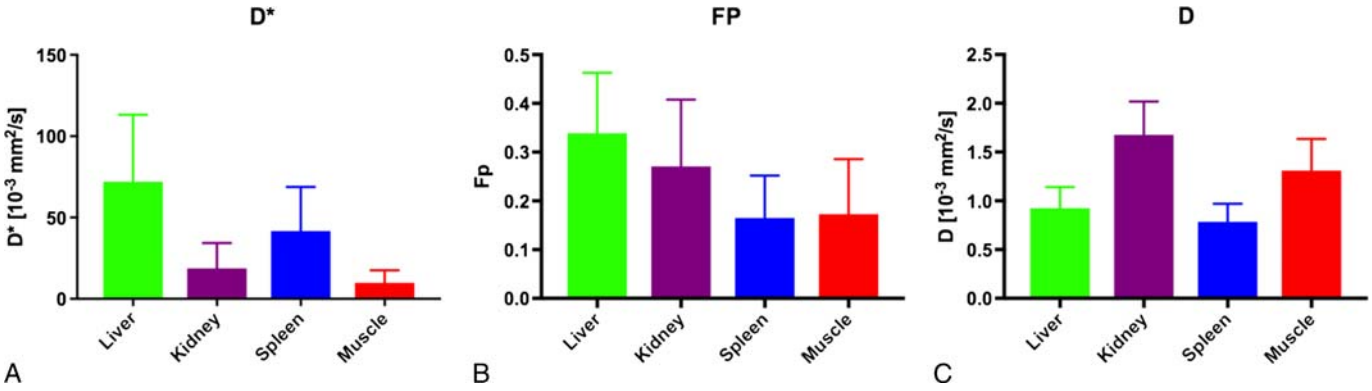

FIGURE 7. IVIM parameters. A, Pseudodiffusion D*. B, Perfusion fraction Fp. C, True diffusion D for the whole cohort. Figure 7 can be viewed online in color at www.investigativeradiology.com.

to the observed shape of the measured IVIM curve, in which, especially, hepatic and, to a lesser extent, splenic parenchyma exhibit a fast signal decline at low b-values, which corresponds to high pseudodiffusion coefficients D* in the IVIM model compared with renal parenchyma and muscle tissue. To differentiate the other tissue types from each other, b-values in the mid and high range had also to be considered. In this respect, b-values ranging from b = 530 s/mm^2 to b = 1000 s/mm^2 were found to be most useful to differentiate splenic and hepatic parenchyma from renal muscle and muscle parenchyma, which corresponds to higher D_t for renal parenchyma and muscle tissue in the IVIM model. This qualitative visual approach was supported by the quantitative approach of PCA.

Thereby, the 16-dimensional variable space was reduced to a 2-dimensional component space via linear combinations of the acquired b-values, assigning each b-value a weighting to one of the both components. The scatter plot of the training data set on the 2-dimensional component space revealed a clear subdivision of the 4 different tissue types, separating hepatic and splenic parenchyma from muscle tissue and renal parenchyma. Taking into account the IVIM parameters and the separation of hepatic and splenic parenchyma from renal parenchyma and muscle tissue via PC2 (*y*-axis) and renal and hepatic parenchyma from muscle tissue and splenic parenchyma via PC1 (*x*-axis), it can be concluded that PC1 reflects the influence of the diffusion coefficient D_t and PC2 the influence of the pseudodiffusion coefficient D*.

Based on the PCA, it was possible to determine quantitatively the 6 b-values, which had most influence on the separation of the different tissue types. Those b-values were in accordance to the previous findings via the scatter plots and covered in particular the low as well as the high spectrum of b-values. A renewed classification of the tissue types with the same number of K-NNs (K = 11 and 13), but applying the reduced set of only 6 b-values, resulted in nearly the same accuracy of 93% to 94% for the training data and even a slightly higher accuracy of 93% for the real-world data set. Besides the fact that the overall performance of the model was very good, with average accuracies above 90% for the training and real-world data set, the classification of hepatic and splenic parenchyma posed more problems than it did with the complete b-value set as both tissue types show a similar behavior for b-values in the low and in the high range. For all applied values of K, hepatic parenchyma was in some cases falsely predicted as splenic parenchyma and vice versa.

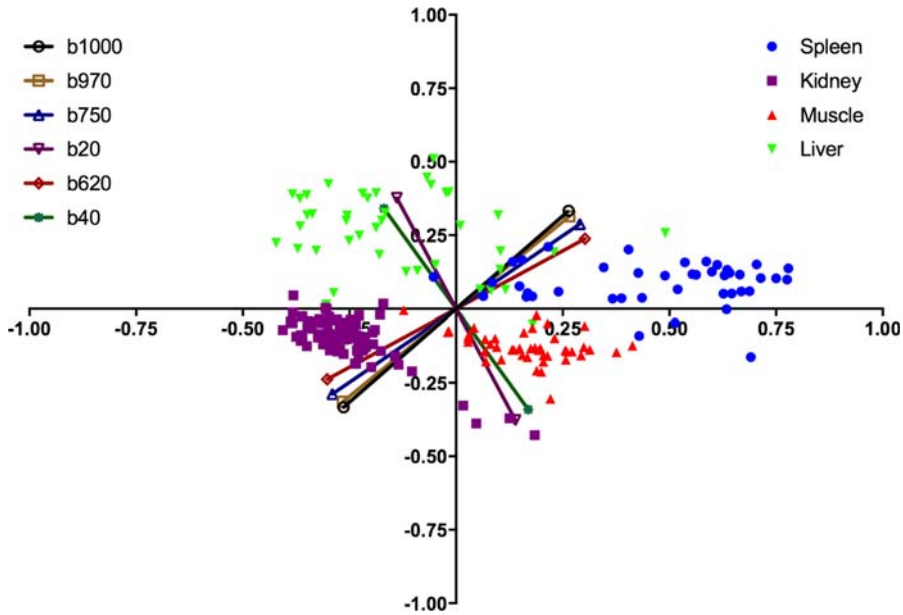

FIGURE 8. Biplot of the implemented PCA, separating all 4 tissue types due to PC1 and PC2. The 6 most important b-values for tissue classification are illustrated by their eigenvectors. Figure 8 can be viewed online in color at www.investigativeradiology.com.

TABLE 5. Eigenvector of Each b-Value Showing the Impact on PC1 and PC2

b-Value, s/mm^2	PC1	PC2	Vector
10	0.116	−0.328	0.348
20	0.139	−0.377	0.402
40	0.169	−0.341	0.381
90	0.221	−0.301	0.373
100	0.230	−0.293	0.372
170	0.265	−0.179	0.320
200	0.282	−0.139	0.314
210	0.292	−0.108	0.311
240	0.287	−0.051	0.291
390	0.313	0.085	0.324
530	0.326	0.188	0.376
620	0.302	0.238	0.385
750	0.290	0.288	0.409
970	0.265	0.314	0.411
1000	0.264	0.333	0.425

PC1, principal component; PC2, second component.

However, the still excellent performance for the reduced b-value data set and the associated low amount of overfitting emphasizes the robustness and stability of the applied machine learning model.

Because the acquisition of a whole IVIM data set is time-consuming and simultaneous multislice scanning (SMS)[24,25] is not yet an integral part in clinical routine, a reduced set of b-values may increase the clinical applicability of IVIM DWI measurements.

To demonstrate the feasibility of IVIM-based tissue classification via machine learning, we only evaluated and used the K-NN algorithm to classify the different tissue types in this study. The reason for this choice is the comprehensibility of the approach, which is more intuitive compared with other machine learning techniques. In future studies, we will evaluate more powerful classifying algorithms such as decision trees, support vector machines, or neural networks.

It may be regarded as a short-coming of the proposed machine learning approach that biomarkers with an intuitive interpretation are missing such as pseudodiffusion corresponding to the degree of tissue perfusion, and true diffusion meaning the restriction of passive water motion due to cell membranes are missing. However, these are concepts that do not make sense to machine learning algorithms, which compare signal values according to a certain scheme. It may be argued that it might be better that the machine learning algorithm decides on its own and determines what patterns are important to predict tissue characteristics such as malignant tumor versus benign tissue alteration. In such a respect, more sophisticated machine learning algorithms such as deep neural networks might be superior compared with the human intuition of biomarkers. In the end, it seems clinically relevant to detect a disease and classify it, whatever biomarkers—intuitive or not—are necessary. In addition, it must be said that, over the past years, many different computational techniques have been described for organ segmentation.[26] This may raise the question, "Why use IVIM data sets for tissue type classification and segmentation?" Because most segmentation techniques are based on surface and landmark detection and typically not relying on organ characteristics except for their shape, these algorithms may run into serious problems for classification, if the anatomy of the organs is severely changed. Whereas in IVIM on the contrary, organs exhibit specific traits, regardless of their actual shape, which can be used as features for automated segmentation with machine learning techniques. Second, IVIM is increasingly used for tissue characterization such as liver fibrosis or treatment monitoring in malignant lesions. Thus, if the DWI sequence is able to perform organ segmentation, these evaluations can be implemented directly on the MR scanner without the need of further postprocessing. Third, in this study, we intended to classify in healthy volunteers different tissue types based on IVIM data as a proof-of-principle test. As in healthy organs, the physiological effects corresponding to specific traits in DWI are mostly well understood; an almost intuitive interpretation of the findings was possible.

Our study exhibited some more limitations. First, the study population was rather small; however, given the fact that we reached a high accuracy on the real-world data set and only low overfitting in the trained model the population size seemed sufficient for this feasibility study. Second, our cohort only consisted of healthy volunteers and we did not include any patients with focal liver lesions or diffusion restrictions to separate healthy from diseased tissue because this will require a much larger study population and will be the topic of future work. Third, different field strengths were not compared. However, we do not expect any differences in the classification of tissue types, because the applied machine learning algorithm was only based on the acquired DW signal intensity differences between the different tissue types.

In conclusion, we have applied a machine learning algorithm with a K-NN model based on IVIM data to classify different tissue types in the abdomen. Thereby, a value of K between 11 and 13 revealed as optimum with respect to the accuracy of the classification and overfitting. By analyzing the data, we were able to identify the b-values, which proved most useful to classify the different tissue types and reduced the complete set of 16 b-values to a set of only 6 b-values, achieving even slightly better accuracy with respect to tissue classification and overfitting. Machine learning has proven capable to classify different tissue types based on IVIM data in a model-free approach without a priori knowledge on physiology.

REFERENCES

1. Bogner W, Gruber S, Pinker K, et al. Diffusion-weighted MR for differentiation of breast lesions at 3.0 T: how does selection of diffusion protocols affect diagnosis? *Radiology*. 2009;253:341–351.
2. Iima M, Yano K, Kataoka M, et al. Quantitative non-Gaussian diffusion and intravoxel incoherent motion magnetic resonance imaging: differentiation of malignant and benign breast lesions. *Invest Radiol*. 2015;50:205–211.
3. Luciani A, Vignaud A, Cavet M, et al. Liver cirrhosis: intravoxel incoherent motion MR imaging–pilot study. *Radiology*. 2008;249:891–899.
4. Donati OF, Chong D, Nanz D, et al. Diffusion-weighted MR imaging of upper abdominal organs: field strength and intervendor variability of apparent diffusion coefficients. *Radiology*. 2014;270:454–463.
5. Schmid-Tannwald C, Oto A, Reiser MF, et al. Diffusion-weighted MRI of the abdomen: current value in clinical routine. *J Magn Reson Imaging*. 2013;37:35–47.
6. Baur AD, Daqqaq T, Wagner M, et al. T2- and diffusion-weighted magnetic resonance imaging at 3 T for the detection of prostate cancer with and without endorectal coil: an intraindividual comparison of image quality and diagnostic performance. *Eur J Radiol*. 2016;85:1075–1084.
7. Barabasch A, Kraemer NA, Ciritsis A, et al. Diagnostic accuracy of diffusion-weighted magnetic resonance imaging versus positron emission tomography/computed tomography for early response assessment of liver metastases to Y90-radioembolization. *Invest Radiol*. 2015;50:409–415.
8. Vallini V, Ortori S, Boraschi P, et al. Staging of pelvic lymph nodes in patients with prostate cancer: usefulness of multiple b value SE-EPI diffusion-weighted imaging on a 3.0 T MR system. *Eur J Radiol Open*. 2015;3:16–21.
9. Yoon JH, Lee JM, Yu MH, et al. Evaluation of hepatic focal lesions using diffusion-weighted MR imaging: comparison of apparent diffusion coefficient and intravoxel incoherent motion-derived parameters. *J Magn Reson Imaging*. 2014;39:276–285.
10. Le Bihan D. Molecular diffusion nuclear magnetic resonance imaging. *Magn Reson Q*. 1991;7:1–30.
11. Le Bihan D, Breton E, Lallemand D, et al. Separation of diffusion and perfusion in intravoxel incoherent motion MR imaging. *Radiology*. 1988;168:497–505.
12. Wurnig MC, Donati OF, Ulbrich E, et al. Systematic analysis of the intravoxel incoherent motion threshold separating perfusion and diffusion effects: proposal of a standardized algorithm. *Magn Reson Med*. 2015;74:1414–1422.

13. Filli L, Wurnig MC, Luechinger R, et al. Whole-body intravoxel incoherent motion imaging. *Eur Radiol*. 2015;25:2049–2058.
14. Andreou A, Koh DM, Collins DJ, et al. Measurement reproducibility of perfusion fraction and pseudodiffusion coefficient derived by intravoxel incoherent motion diffusion-weighted MR imaging in normal liver and metastases. *Eur Radiol*. 2013;23:428–434.
15. Ichikawa S, Motosugi U, Ichikawa T, et al. Intravoxel incoherent motion imaging of the kidney: alterations in diffusion and perfusion in patients with renal dysfunction. *Magn Reson Imaging*. 2013;31:414–417.
16. Chandarana H, Kang SK, Wong S, et al. Diffusion-weighted intravoxel incoherent motion imaging of renal tumors with histopathologic correlation. *Invest Radiol*. 2012;47:688–696.
17. Schneider MJ, Dietrich O, Ingrisch M, et al. Intravoxel incoherent motion magnetic resonance imaging in partially nephrectomized kidneys. *Invest Radiol*. 2016; 51:323–330.
18. Cohen AD, Schieke MC, Hohenwalter MD, et al. The effect of low b-values on the intravoxel incoherent motion derived pseudodiffusion parameter in liver. *Magn Reson Med*. 2015;73:306–311.
19. Lemke A, Stieltjes B, Schad LR, et al. Toward an optimal distribution of b values for intravoxel incoherent motion imaging. *Magn Reson Imaging*. 2011;29:766–776.
20. Pang Y, Turkbey B, Bernardo M, et al. Intravoxel incoherent motion MR imaging for prostate cancer: an evaluation of perfusion fraction and diffusion coefficient derived from different b-value combinations. *Magn Reson Med*. 2013; 69:553–562.
21. Cover TM, Hart PE. Nearest neighbor pattern classification. *IEEE Trans Inform Theory*. 1967;IT-13:21–27.
22. Moore WA, Khatri G, Madhuranthakam AJ, et al. Added value of diffusion-weighted acquisitions in MRI of the abdomen and pelvis. *AJR Am J Roentgenol*. 2014;202:995–1006.
23. Klauss M, Lemke A, Grünberg K, et al. Intravoxel incoherent motion MRI for the differentiation between mass forming chronic pancreatitis and pancreatic carcinoma. *Invest Radiol*. 2011;46:57–63.
24. Obele CC, Glielmi C, Ream J, et al. Simultaneous multislice accelerated free-breathing diffusion-weighted imaging of the liver at 3T. *Abdom Imaging*. 2015; 40:2323–2330.
25. Filli L, Ghafoor S, Kenkel D, et al. Simultaneous multi-slice readout-segmented echo planar imaging for accelerated diffusion-weighted imaging of the breast. *Eur J Radiol*. 2016;85:274–278.
26. Norajitra T, Maier-Hein KH. 3D statistical shape models incorporating landmark-wise random regression forests for omni-directional landmark detection. *IEEE Trans Med Imaging*. 2017;36:155–168.